B24 E 2-?

FACULTY WORKING PAPER 92-0154

Using Inductive Learning to Predict Bankruptcy

James A. Gentry
Department of Finance
University of Illinois

Michael J. Shaw
Department of Business Administration
University of Illinois

Antoinette C. Tessmer
Department of Finance
University of Illinois

David T. Whitford
Department of Finance
University of Illinois

Bureau of Economic and Business Research
College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 92-0154

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

August 1992

Using Inductive Learning to Predict Bankruptcy

James A Gentry
Department of Finance

Michael J. Shaw
Department of Business Administration

Antoinette C. Tessmer
Department of Finance

David T. Whitford
Department of Finance

USING INDUCTIVE LEARNING TO PREDICT BANKRUPTCY

James A. Gentry
IBE Distinguished Professor of Finance
University of Illinois, Urbana-Champaign

Michael J. Shaw
Associate Professor of Business Administration
University of Illinois, Urbana-Champaign

Antoinette C. Tessmer
Visiting Assistant Professor of Finance
University of Illinois, Urbana-Champaign

David T. Whitford
Associate Professor of Finance
University of Illinois, Urbana-Champaign

ABSTRACT

The primary contribution of this study is the use of cash flow components in an inductive learning system to predict financial failure. The underlying conceptual framework associated with inductive learning is presented, and an example of entropy is developed in an appendix. The sample included 14 cash flow components and two qualitative variables for 198 companies, 99 failed and 99 nonfailed companies. These inputs were used in a C4.5 inductive learning program to predict the failed/nonfailed status of the sample companies. The program induces a decision tree that reflects the structure of the inputs used to classify the companies as being failed or nonfailed. A global tree interpretation combined with a jackknife procedure was used to repeat the experiment 198 times, which resulted in a predictive accuracy of 86 percent. A global tree represents a composite of the 198 induced trees. The global tree approach indicates that knowing the level of dividends (DIV^*), net capital investment (NIF^*) and net operating cash flow (NOF^*) results in the correct identification of 89 percent of the failed and nonfailed companies. The inductive learning test results were superior to the 67.5 percent predictive accuracy of a series of probit tests. The results of these tests are encouraging and indicate the need for further study in the use of inductive learning systems in predicting and interpreting financial performance.

USING INDUCTIVE LEARNING TO PREDICT BANKRUPTCY[1]

Since the mid 1960s numerous empirical models have been developed that use annual financial information to discriminate between firms that declare bankruptcy and the ones that remain solvent.[2] In general these models lack an underlying theory (Scott [1976]) and their results are dependent on the time period studied, the firms included in the sample and the statistical methodology used. During the past decade a separate stream of studies used market determined returns and risk measures to explain the bankruptcy process, reorganization, and the costs associated with bankruptcy.[3] Finally, there was a third stream of theoretical research that used security pricing formulas to explain corporate bankruptcy.[4]

Although the bankruptcy literature is extensive, there is continued interest in the development of a theoretical foundation that would capture the many dimensions of financial distress and failure. Likewise there are numerous lenders and investors who are deeply interested in improving their ability to explain, interpret and predict bankruptcy. Most of the studies use financial ratios in a statistical model such as multiple discriminant analysis, probit or logit. However, cash flow information provides unique and subtle insights into the prediction of bankruptcy, bond ratings and loan risk ratings.[5] A fundamental contribution of this study is to use cash flow components in an inductive learning system to predict if a firm is either bankrupt or nonbankrupt. Inductive learning is a relatively new analytical approach that is based on an information theory concept called entropy.

This paper is organized in the following manner. The next section briefly reviews the calculation of the cash flow components. It presents a hierarchy of cash flow components and provides a theoretical explanation of using these components to interpret financial strengths and weaknesses. Section III provides an explanation of the inductive learning system used to predict bankruptcy. The sample used to test the model is found in Section IV. An interpretation of the decision tree generated by the inductive learning system is presented in Section V and conclusions associated with this study are in Section VI.

II. CASH FLOW COMPONENTS

Gentry, Newbold, and Whitford [1985, 1990] developed a total cash flow system with 12 cash flow components (CFC). The objective was to integrate cash flow information from the income statement and the balance sheet, i.e., changes in the items between two periods. The total cash flow system provides unique insight concerning management's allocation of resources and the overall performance of the firm.[6] An example of the 12 CFC are presented at the top of Exhibit 1.

A relative cash flow component (CFC^*) represents the percentage contribution of each CFC to the total cash flow. A relative cash flow component is determined by dividing each component by the total cash flow (CFC/TCF). An example of CFC^* are presented at the bottom of Exhibit 1. A brief overview shows the proportion each component contributes to the total cash flow. Exhibit 1 shows that 59.8% of the total inflow came from operations, 16.7% was from net financing, and 9.8% from payables. On the outflow side, which are identified with a

minus (-) sign, net investment represented 35.3% of the total outflow, receivables composed 21.6%, inventories 17.6%, and dividends 14.7%.

The CFC^* in Exhibit 2 are arranged in a hierarchical order that reflects their economic importance in evaluating the financial health of a firm. Generally, financial and credit analysts use the hypothesized cash flow hierarchy to evaluate a firm's financial strengths and weaknesses. The CFC^* hierarchical structure highlights the contribution each component makes to the net cash flow surplus or deficit. An example of the CFC^* hierarchy and the relative net cash flow (NCF^*), i.e., the net surplus or deficit cash flow position, is presented in Exhibit 2. This example is based on research findings of Gentry, Newbold, and Whitford [1990].

By definition Company A has the lowest credit risk, which is based on the composition of the CFC^*. Exhibit 2 shows 92% of Company A's cash inflows originate from operations (NOF^*), which is the highest NOF^* among the four credit risk classes. After deducting from NOF^* the major outflows for investment--NIF^* (-45%), the highest among the four credit risk classes, and changes in net working capital (-13%), the remaining cash flow surplus represents 34% of the total. The 34% surplus is the highest among the four credit risk classes. The two major outflows associated with the costs of external financial capital are interest expense [fixed coverage expenditures (FCE^*)] and dividends (DIV^*). After deducting the FCE^*, which is the lowest among the four credit risk classes, the surplus cash flow available for dividends (DIV^*) is 32%. DIV^* consume 12% of total outflows, which leaves a net cash flow surplus

of 20%. The surplus cash is used to retire debt (-10%) and invest in marketable securities (-10%).

In contrast Company D is an example of a distressed company and it is in the highest credit risk class. Company D has 15% of its cash inflows coming from operations, which is the lowest NOF* among the four risk classes. After deducting cash outflows of 18% for total investment, NIF* being 15% and a net reduction in working capital is 3%, Company D has a deficit cash flow equal to -3% of the total cash flow. The cash outflow to NIF* and networking capital is the smallest among the four credit risk classes. The FCE* represents 16% of the total outflow, which leaves a -19% to pay DIV*. The interest payment for Company D is the largest among the four credit risk classes and the deficit before DIV* is also the largest. DIV* adds an additional 1% to total outflow, the lowest among the four groups. The -20% represents a net cash flow deficit and shows that Company D has used all of its operating and working capital cash inflows plus an additional 20% to cover the outflows for investment, dividends and fixed coverage expenditures. Exhibit 2 also shows the deficit was offset by an increase in financing ΔNF equals 19%, and a decrease in net other assets and liabilities of 1%.

Exhibit 2 illustrates several basic concepts that exist between the net cash flow surplus or deficit and levels of risk. First, as the percentage of cash inflows from net operations declines, the net cash flow surplus becomes smaller or alternatively, a deficit becomes larger. Second, as the net cash flow surplus declines or the net cash flow deficit increases, a firm's financial risk increases. For example,

of 20%. The surplus cash is used to retire debt (-10%) and invest in marketable securities (-10%).

In contrast Company D is an example of a distressed company and it is in the highest credit risk class. Company D has 15% of its cash inflows coming from operations, which is the lowest NOF* among the four risk classes. After deducting cash outflows of 18% for total investment, NIF* being 15% and a net reduction in working capital is 3%, Company D has a deficit cash flow equal to -3% of the total cash flow. The cash outflow to NIF* and networking capital is the smallest among the four credit risk classes. The FCE* represents 16% of the total outflow, which leaves a -19% to pay DIV*. The interest payment for Company D is the largest among the four credit risk classes and the deficit before DIV* is also the largest. DIV* adds an additional 1% to total outflow, the lowest among the four groups. The -20% represents a net cash flow deficit and shows that Company D has used all of its operating and working capital cash inflows plus an additional 20% to cover the outflows for investment, dividends and fixed coverage expenditures. Exhibit 2 also shows the deficit was offset by an increase in financing ΔNF equals 22%, and a decrease in net other assets and liabilities of 3%.

Exhibit 2 illustrates several basic concepts that exist between the net cash flow surplus or deficit and levels of risk. First, as the percentage of cash inflows from net operations declines, the net cash flow surplus becomes smaller or alternatively, a deficit becomes larger. Second, as the net cash flow surplus declines or the net cash flow deficit increases, a firm's financial risk increases. For example,

Firm A has the highest net cash flow surplus and it has the lowest financial risk. In contrast, Firm D has the largest net cash flow deficit and it has the highest financial risk. Third, as the relative cash inflow from operations (NOF^*) decreases, the relative cash outflow to capital investment decreases. That is the percentage of cash outflow going to investment is closely related to operating cash inflows. In turn, as the relative cash outflow for interest expense (FCE^*) increases, the outflow for DIV^* decreases. Furthermore, the trend of FCE^* is negatively related to NOF^* and NIF^*. The pattern of the interrelationships among the key cash flow components is closely associated with the financial health of a firm.[7]

III. THE ID3 METHOD: INDUCTION OF DECISION TREES

ID3, Quinlan [1986], is an inductive learning program based on the original work of Hunt [1966]. Using data cases pertaining to a known class and described in terms of a fixed set of attributes, ID3 produces a decision tree based on the attributes that correctly classify the given cases. The induction of a decision tree is based on the process of splitting a group of training examples according to the value of a selected attribute, where the examples in a group belong to the same class. Thus, an important step in building the decision tree is selecting the best attribute to branch. ID3 employs a measure of <u>entropy</u> as a yardstick for this selection.

The concept of entropy originated in the field of the natural sciences, Halliday [1978], and was later used in the field of information sciences, Shannon [1948, 1951]. Thermodynamics theories contend that when the entropy of a system tends to increase, the

disorder of this system tends to increase as well, Halliday [1978]. In the theories related to communication and psychology, the same concept is used to measure the amount of randomness or uncertainty contained in a message. Suppose a message consists of an event with two possible outcomes, x and x', with probabilities p and (1-p) to occur. The uncertainty about which outcome will actually be encountered is calculated as the entropy of that message:

$$H(x) = -\sum_{i=1}^{n} q(x_i) \log_2 q(x_i)$$

which is reduced to

$$H = -p \log_2 p - (1-p) \log_2 (1-p)$$

for the case of an event with two possible outcomes.

When p = 0 or 1, there is no uncertainty about the outcome of the event and hence the entropy equals zero, H = 0. When p = 1/2, there exists maximum uncertainty as to whether x or x' will occur, and hence H has the maximum value, as shown in Figure 1. Therefore, the higher the entropy (H), the more <u>uncertainty</u> about the content of the message, Ash [1965].

Shannon uses the entropy measure in his attempt to solve one of the fundamental problems of communication: to reproduce at one point either exactly or approximately a message transmitted from another point, via a discrete channel for transmitting information, e.g.,

teletype or telegraph. Entropy, used as a measure of the uncertainty contained in alternative possible messages, helps to select the best reproduction of the incoming message, Shannon [1963]. The discrete channel for transmitting information is used to reduce the uncertainty contained in the incoming message and to produce an outgoing message containing the least uncertainty. In the theory of communication, information is also defined as that which removes or reduces uncertainty, Attneave [1959]. Thus, information and entropy appear as closely related concepts: the amount of information is determined by the amount that uncertainty is reduced. The entropy measure gives therefore, by complement, a measure of the amount of information contained in a message.

In ID3, the decision tree for classifying data cases may be regarded as Shannon's channel for transmitting information that produces a message indicating the classification for a given data case. When a node of the tree contains only data cases of the same class, the entropy of the message associated with that node is equal to zero, which means that the classification decision is certain and defined for the data cases belonging to that node. The induction of the decision tree is thus the process of selecting an attribute to branch that results in the maximal reduction of entropy--which can also be viewed as a process of maximizing information gains.

Starting with a root node, the decision tree is generated by selecting progressively attributes to branch the tree. At each iteration of generating the tree, ID3 examines all candidate attributes and chooses the attribute that can reduce the amount of entropy

contained in the current version of the decision tree. In other words, ID3 chooses the attribute that maximizes the amount of information gained. This process is illustrated in Appendix A.

ID3 follows a top-down, divide-and-conquer approach for specializing during the process of induction, i.e., the process subdivides and assigns the cases of the training set at a node into two or more smaller subsets. Therefore, the longer the tree, the more it is specialized to specific cases subsets. Consequently, generalization of a decision tree, which is the inverse of specialization, can be achieved by pruning the tree from the bottom-up based on some evaluating criterion. This is the case for the C4.5 version of ID3 program used in this study.

Examples of the criteria that are used are: (1) the complexity of the resulting tree, (2) the number of terminal nodes in the tree, Breiman, et al., [1984], and (3) the number of instances present at a node that represent each of the classes. The last case occurs because the number of instances decreases as we traverse along a branch of a decision tree from top to bottom, which leads to insignificant splitting due to inadequate sample sizes. In reducing the complexity of decision trees by pruning, Breiman, et al. [1984] used the number of terminal nodes and the misclassification cost of the generated tree as a measure of computational complexity.

Pruning not only reduces the size of a decision tree, it decreases the effect of noise in the data. Real-world data used in a training example contain a reasonable amount of noise. The negative effect of noise increases from the root of the tree downward because the terminal

nodes contain a smaller number of cases per represented class. Pruning helps to reduce the propagation of the error by maintaining the number of cases per class at any given node at a desired level. Consequently, pruning reduces the effect of noise. Pruning a tree may increase the number of classification errors made on the training data, but should decrease the error rate on the independent test data, Mingers [1989, p. 228].

IV. DATA

To be included in the sample, each company had to have complete annual balance sheet and income statements that were released for the two fiscal years prior to the date that the bankruptcy was declared. This insured that the financial statements were available to the public. The source of the data was the Compustat PC Plus database, the sample criteria resulted in 106 industrial firms that had declared bankruptcy or had been liquidated during the period 1971-1987.

The 106 bankrupt companies were matched with a company that had a similar 4-digit SIC code and comparable annual sales for the year immediately prior to the bankruptcy declaration. Three companies were eliminated from the database because a matching company was unavailable. Finally four companies were eliminated because of incomplete data for the matching firms. The final sample was composed of 99 failed companies, that had a dummy variable value of 1, and 99 nonfailed companies with a dummy variable of 0, for a total of 198 sample companies.

A holdout sample of 40 failed and 40 nonfailed companies were randomly selected from the total sample. To test the stability of the

inductive learning model, a total of five holdout samples were randomly selected. The five training samples were composed of the remaining 59 failed companies and the 59 matching nonfailed companies in each respective sample.

The training sample contained 11 relative cash flow variables. In addition there were three other variables included in the training set: the first variable, total cash flow divided by total assets (TCF/TA), was included for scalar purposes. Additionally, two qualitative measures were included. It is hypothesized that older assets are less efficient than newer assets and firms with older assets are more likely to experience financial failure. The age of the assets employed by the firm is calculated by dividing accumulated depreciation by the historical cost of the fixed assets, that is, Accumulated $\text{Depreciation}_t/\text{Fixed Assets}_t$. The second qualitative variable determines the trend of sales during the year before bankruptcy was declared. If the sales trend was upward during the year before bankruptcy, a dummy variable was assigned a value of zero. If the sales trend was downward, the company was assigned a value of one.

V. INDUCTIVE LEARNING ANALYSIS

The balance sheet and income statement information for the 118 companies was used to determine the cash flow components for 59 failed companies and 59 nonfailed companies. The means and standard deviations for each of the 12 cash flow components, TCF/TA and the two qualitative variables are presented in Exhibit 3.

The inductive learning approach is based on the training examples to learn a structure of the decision-making process. The structure

determined by the training example is then used to test a holdout sample referred to as the testing sample. The information used in the training example is the 11 relative cash flow components, TCF/TA and the two qualitative measures. The C4.5 inductive learning system uses these 14 variables to predict the failed or nonfailed status of each training company. The entropy method selects the variables according to the amount of information added at each level of the decision tree as shown in Appendix A.

A hierarchy of the relative cash flow components (CFC^*) is presented in Exhibit 2. The structure of the cash flow hierarchy establishes a theoretical foundation for hypothesizing the structure of a decision tree generated by the C4.5 system. That is, the net operating cash flow (NOF^*) would be the root node followed closely, but not in any specific order, by net investment (NIF^*), dividends (DIV^*), fixed coverage expenditures (FCE^*) and the five working capital variables (ΔARF^*, $\Delta INVF^*$, $\Delta OCAF^*$, ΔAPF^*, $\Delta OCLF^*$). We do not have a theory to hypothesize where TCF/TA and the qualitative variables will appear in the structure.

In testing the accuracy and stability of the C4.5 inductive learning system, initially five separate trees were generated. Each tree had a unique structure and used a different combination of attributes. The decision tree in Figure 2 is presented as a reasonable proxy of the five trees generated by C4.5. A brief explanation of the decision tree helps interpret the structure of the quantitative and qualitative variables generated by C4.5. Among the 14 variables in the

training set, the inductive learning process found DIV^* to be the most discriminating variable, i.e., DIV^* is the root node of the tree.

Figure 2 shows that 67.8 percent (40/59) of the failed companies are correctly classified by knowing that dividends are very close to zero, i.e., the proportion of cash outflow going to dividends is greater than -.001, which is zero. The remaining 78 companies (118-40) disbursed more than .1 percent of their total cash outflows to dividends.

At the second node there were five companies for which net investment (NIF^*) represented more than 1.14 percent of their total cash inflows. These five companies were correctly classified as failed firms. The remaining 73 companies had a NIF^* of less than 1.14 percent, which for most of the remaining companies reflects a cash outflow for capital expenditures. Thus the C4.5 system has selected two cash flow variables, DIV^* and NIF^*, which resulted in approximately 76 percent (45/59) of the failed training companies being classified correctly.

The inductive learning system found the net financing flow component (ΔNFF^*) to be the third most important variable in classifying failed and nonfailed companies. Figure 2 shows 11 companies that used cash to retire debt or equity were classified as being nonfailed companies. Ten of these companies, which had a ΔNFF^* of less than -17 percent, were correctly classified, but the eleventh firm was incorrectly classified.

At the fourth level nine companies with an accumulated depreciation/total fixed asset ratio of less than 22.14 percent were correctly classified as nonfailed firms. This is the first qualitative

variable to be selected by the C4.5 system. At the fifth level two companies with an accumulated depreciation/total fixed asset ratio between 22.14 percent and 25.23 percent were correctly classified as failed companies.

Accounts payable was selected by the inductive learning system as the most discriminating variable at the sixth level. Twenty three companies whose accounts payable represented at least 2.85 percent of their total cash inflow were classified as nonfailed companies. The C4.5 system correctly classified 22 of these companies, but one was incorrectly classified. A sequential linear pattern existed in the selection of the first six levels of information. At the sixth level almost 80 percent of the failed companies and nearly 70 percent of the nonfailed companies have been correctly classified.

At the seventh level 5.61 percent or more of cash outflow going to dividends correctly classified 12 companies as nonfailed and misclassified two firms. Finally, net other assets and liabilities (ΔNOA&LF) was the eighth variable needed to classify 10 companies as failed and four companies as nonfailed.

In summary the decision tree in Figure 2 shows that the inductive learning system correctly classified 96.4 percent (114/118) of the companies in the training sample. This pattern of data created from the training set is used to predict the failed/nonfailed status of 80 companies in the separate testing sample. The inductive learning system correctly predicted the failed/nonfailed status of 90.2 percent (72/80) of the testing sample.

Global Tree Interpretation

Using a single tree to represent a common structure of the data presents a challenge to the credit analysts. Each training data set produces a unique structure that has a different combination of attributes. The C4.5 system reduces the complexity of decision trees by pruning, Breiman, et al. [1984]. Pruning reduces the size of a decision tree and decreases the effect of noise in the real world data. However, it does not help stabilize the structure of the tree.

The challenge is to find a common structure that reflects stability in the horizontal location of the variables, as well as vertical stability associated with the length of the tree. A global tree interpretation process developed by Tessmer [1992] uses a jackknife procedure to develop a model of failure prediction. The first step is to use the C4.5 system to induce a set of original decision trees. Because each induced tree can have a unique structure, Tessmer [1992, pp. 12-15], the jackknife procedure was used to repeat the experiment 198 times. The jackknife procedure resulted in a mean predictive accuracy of 86 percent.

The global tree interpretation resulted in the creation of a final global tree shown in Figure 3, Tessmer [1992, pp. 12-15]. The final global tree is a composite of the 198 original trees and contains attributes that appeared in 50 percent or more of the decision trees induced by the C4.5 program. The global tree reduces noise and overfitting effects that are present in the original trees. Figure 3 retains the most frequently appearing attributes in their most likely position in the original trees.

The global tree shown in Figure 3 needed only three attributes--dividends (DIV^*), net investment (NIF^*), and net operating cash flows (NOF^*)--to classify the 198 companies as being either failed or nonfailed. Figure 3 shows on average the global tree with three cash flow attributes correctly classified 88.9 percent (176/198) of the failed and nonfailed companies. That is the inductive learning system correctly classified 83.8 percent (83/99) of the failed companies and 93.9 percent (93/99) of the nonfailed companies.

The root node of the global tree was the dividend (DIV^*) component. By knowing that a company did not pay a dividend, C4.5 correctly classified 70 percent (69/99) of the failed companies. A three dimension frequency diagram of DIV^* for the failed and nonfailed companies is presented in Figures 4 and 5, respectively. These two figures highlight why DIV^* was selected as the root node, the most discriminating attribute. Figure 4 shows nearly 70 percent of the failed companies had a DIV^* component that ranged from zero to 5 percent. Figure 2 indicates that for 40 of the 70 companies DIV^* was zero. The DIV^* component for the remaining failed companies is scattered across a range from -5 percent to -45 percent. Figure 5 shows the DIV^* for the nonfailed companies was widely disbursed across a range from zero to 35 percent. Thus in contrast to the failed companies, the DIV^* component of the nonfailed companies is not heavily concentrated in a single cell.

Another 10 percent of the failed companies are correctly classified by learning that capital investment (NIF^*) was a cash inflow. Finally, knowing that DIV^* and NIF^* were cash outflows greater than zero

and learning that net operating cash flows (NOF^*) was positive, i.e., greater than zero, 94 percent (93/99) of the nonfailed companies were correctly identified. Also learning that NOF^* was negative made it possible to identify an additional four failed companies.

Focusing Observations

Several significant observations evolve from the analysis. Initially, it was hypothesized that the net operating cash flow component (NOF^*) would be the root node in the induced decision trees. However, the inductive learning results show that DIV^* was the root node, that is the most discriminating cash flow component in classifying loan risk. This finding supports previous empirical test results that predicted bond ratings and bankruptcy, Gentry, Newbold and Whitford [1985a, 1985b, 1988]. Why isn't NOF^* the root node as hypothesized? It is our interpretation that DIV^* is a proxy for NOF^*. The surplus cash flow available for paying dividends is dependent on a firm's operating performance in the execution of its strategic plans. Although there are several decisions and actions responsible for generating a surplus net cash flow, NOF^* is the theoretical foundation for creating a surplus cash flow that can be used to pay dividends. In essence, DIV^* reflects a firm's dividend policy, but more importantly it provides a signal to the financial markets that the firm has the cash available to pay dividends to its shareholders.[8]

Tree induction reveals several characteristics of the cash flow data being analyzed. First, the presence of only a few nodes on the tree signals that distinct information patterns exist which make it possible to discriminate between failed and nonfailed companies.

Second, a small linear tree indicates that a straight sequence of a few attributes can easily determine the failed/nonfailed status of a firm. Third, the most discriminating and important attributes are close to the root node. Fourth, the value added by the components in the lower levels of the tree is markedly less than the value contributed by the components closer to the root of the tree. If several components are needed to determine a firm's fail/nonfail status, it indicates across firms there is complex and noisy information that makes it difficult to differentiate between failed and nonfailed companies.

Probit Analysis

A final set of tests were undertaken to provide further insight into the above results. The same five data sets were used to develop probit models to classify and predict the failed/nonfailed status of the sample companies. On average these probit models correctly classified 81.4 percent (96/118) of the companies in the training sample. Two variables were statistically significant at the .01 level of significance--dividends and net investment. The coefficients of these probit models were used to predict the failed/nonfailed status of their holdout samples. The predictive test results correctly identified on average the status of 67.5 percent (54/80) of the companies in the holdout samples. Both test results are quite acceptable, but in this experiment the inductive learning model produced superior results.

VI. CONCLUSIONS

The objectives of this paper were to use cash flow components and qualitative variables in an inductive learning system to predict

financial failure. One of the primary advantages of an inductive learning system is the insightful decision structure it provides for interpreting financial performance. Each sample data set generated by the inductive learning system produces a unique structure that has a different combination of attributes. To determine if there was stability in the structure a global tree interpretation was introduced into the analysis. A jackknife procedure was used to repeat the experiment 198 times which resulted in a predictive accuracy of 86 percent. Furthermore, the global tree procedure developed a composite of the 197 induced trees, and indicated that knowing the level of dividends (DIV^*), net capital investment (NIF^*) and net operating cash flows (NOF^*) resulted in the correct identification of 89 percent of the failed and nonfailed companies. A probit model produced a 67.5 percent predictive accuracy. In conclusion, using cash flow components in an inductive learning system provided a high level of predictive accuracy. Also it selected attributes that closely resembled a hypothesized hierarchical structure of the cash flow components.

Footnotes

[1]The authors are very appreciative of the financial support provided by the KPMG Peat Marwick Foundation for sponsoring this research project. The authors are grateful for the very capable research assistance of Brian Bielinski and Joe Deters.

[2]For example, Altman [1968], Altman, Haldeman and Narayaman [1977], Ball and Foster [1982], Beaver [1966], Casey and Bartczak [1985], Gentry, Newbold and Whitford [1985a, 1985b], Lane, Looney and Wansley [1986], Ohlsen [1980], and Zmijewski [1984].

[3]For example, Aharony, Jones and Swary [1980], Betker [1990], Clark and Weinstein [1983], Franks and Torous [1989, 1990], Haugen and Senbet [1978, 1988], Morse and Shaw [1988], Warner [1977a, 1977b], and White [1983].

[4]For example, Black and Scholes [1973], Scott [1976, 1981], and Stiglitz [1972].

[5]See for example Gentry, Newbold and Whitford [1985a, 1985b, 1987, 1988 and 1991].

[6]Recent studies examined the incremental information content of cash flows given earnings, Wilson [1986, 1987], Bowen, et al. [1987] and Rayburn [1986], and generally found the existence of information content in cash-flow data. Bernard and Stober [1989] disaggregated net income and found it did not provide additional information content beyond net income. Livnat and Zarowin [1990] examined the components of cash flows from financing, investing and operating activities for differential associations with annual security returns.

[7]In evaluating the strategic performance of companies, Donaldson [1984] developed a model for measuring sustainable growth. The model was based on two variables--the rate of growth of sales (gS) and the rate of return on net assets (RONA). If the rate of growth of sales exceeded the rate of returns on net assets, gS > RONA, the firm experienced a deficit cash flow. Such a finding indicates the firm was not generating sufficient cash flow to sustain its future growth, e.g., Company's C and D in Exhibit 1. However, if RONA > gS, the firm had surplus cash flow, e.g., Firm A in Exhibit 1. Under these circumstances, the firm could sustain a higher rate of growth of sales if acceptable investment alternatives were available. Frequently, large firms with relatively mature product lines experience surplus cash flow, e.g., Company A in Exhibit 1. Finally, Donaldson observed that firms strive for an annual cash flow that approaches zero. That is, where gS = RONA, which allows the firm to meet its investment schedule without having to use the capital markets, e.g., Company B in Exhibit 1.

[8]Miller and Rock [1985, p. 1046] observed the best places to look for signalling may well be among firms falling into adversity, not because they start signalling but because they stop.

References

Aharony, J., C. P. Jones, I. Swary, "An Analysis of Risk and Return Characteristics of Corporate Bankruptcy Using Capital Market Data," _Journal of Finance_, Vol. 35 (September 1980), pp. 1001-1016.

Altman, E. I., "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," _Journal of Finance_, Vol. 23 (September 1968), pp. 589-609.

____________, R. Haldeman and P. Narayaman, "ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations," _Journal of Banking and Finance_, Vol. 1 (June 1977), pp. 29-54.

Attneave, F., _Applications of Information Theory to Psychology: A Summary of Basic Concepts, Methods, and Results_, Holt, Rinehart and Winston, New York, NY, 1959.

Ball, R. and G. Foster, "Corporate Financial Reporting: A Methodological Review of Empirical Research," _Journal of Accounting Research_, Vol. 20 (Supplement 1982), pp. 29-54.

Beaver, W., "Financial Ratios as Predictors of Failure," _Journal of Accounting Research_, Vol. 4 (Supplement 1966), pp. 71-102.

Bernard, V. L. and T. L. Stober, "The Nature and Amount of Information in Cash Flows and Accruals," _The Accounting Review_, Vol. 65 (October 1989), pp. 624-652.

Betker, B., "An Analysis of the Returns to Stockholders and Bondholders in a Chapter 11 Reorganization," UCLA Working Paper, November 20, 1990, 43 pages.

Black, F. and M. Scholes, "The Pricing of Options and Corporate Liabilities," _Journal of Political Economy_, Vol. (May-June 1973), pp. .

Bowen, R. M., D. Burgstahler and L. A. Daley, "The Incremental Information Content of Accrual versus Cash Flows," _The Accounting Review_, Vol. 62 (October 1987), pp. 723-747.

Breiman, L., Freidman, J. H., Olshen, R. A., and Stone, C. J., _Classification and Regression Trees_, California: Wadsworth Publishing Company, 1984.

Casey, C. and N. Bartczak, "Using Operating Cash Flow Data to Predict Financial Distress: Some Extensions," Journal of Accounting Research, Vol. 23 (Spring 1985), pp. 384-401.

Clark, T. and M. I. Weinstein, "The Behavior of the Common Stock of Bankrupt Firms," Journal of Finance, Vol. 38 (May 1983), pp. 489-504.

Donaldson, G., Managing Corporate Wealth: The Operation of a Comprehensive Financial Goals System, New York: Praeger Publishers, 1984.

Franks, J. and W. N. Torous, "An Empirical Investigation of U.S. Firms in Reorganization," Journal of Finance, Vol. 44 (July 1989), pp. 747-769.

Gentry, J., P. Newbold, and D. T. Whitford, "Classifying Bankrupt Firms with Funds Flow Components," Journal of Accountancy Research, Vol. 23 (Spring 1985a), pp. 146-160.

____________, "If Cash Flow's Not the Bottom Line, What Is?" Financial Analysts Journal, Vol. 41 (September/October 1985b), pp. 47-56.

____________, "Funds Flow Components, Financial Ratios and Bankruptcy," Journal of Business Finance and Accountancy, Vol. 14 (Winter 1987), pp. 595-606.

____________, "Predicting Industrial Bond Ratings with a Probit Model and Funds Flow Components," Financial Review, (August 1988), pp. 269-286.

Gentry, J., M. J. Shaw and D. T. Whitford, "Predicting Loan Risk with Systems that Use an Inductive Learning Approach," Working Paper, University of Illinois, March 1991, 35 pages.

Halliday, D. and R. Resnick, Physics-Part 1 and 2 (Combined 3rd Edition), John Wiley & Sons, New York, NY, 1978.

Haugen, R. A. and L. W. Senbet, "The Insignificance of Bankruptcy Costs to the Theory of Optimal Capital Structure," Journal of Finance, Vol. 33 (May 1978), pp. 383-393.

____________, "Bankruptcy and Agency Costs: Their Significance to the Theory of Optimal Capital Structure," Journal of Financial and Quantitative Analysis, Vol. 23 (March 1988), pp. 27-38.

Hunt, E. B., Concept Learning: An Information Processing Problem, New York, NY: John Wiley & Sons, 1962.

Jensen, M. C., "The Takeover Controversy: Analysis and Evidence," Midland Corporate Finance Journal, Vol. 4 (Summer 1986), pp. 6-32.

Jensen, M. C. and R. S. Ruback, "The Market for Corporate Control: The Scientific Evidence," Journal of Financial Economics, Vol. 11 (April 1983), pp. 5-50.

Kaplan, R. and G. Urwitz, "Statistical Model of Bank Ratings: A Methodological Inquiry," Journal of Business, Vol. 52 (April 1979), pp. 231-261.

Lane, W. R., W. Looney and J. W. Wansley, "An Application of the Cox Proportional Hazard Model to Bank Failure," Journal of Banking and Finance, Vol. 10 (1986), pp. 511-531.

Lehn, K. and A. Poulsen, "Free Cash Flow and Stockholder Gains in Going Private Transactions," Journal of Finance, Vol. 44 (July 1989), pp. 771-787.

Livnat, J. and P. Zarowin, "The Incremental Information Content of Cash-Flow Components," Journal of Accounting and Economics, Vol. 3 (1990), pp. 25-46.

Miller, M. and K. Rock, "Dividend Policy Under Asymmetric Information," Journal of Finance, Vol. 40 (September 1985), pp. 1031-1051.

Mingers, J., "An Empirical Comparison of Pruning Methods for Decision Tree Induction," Machine Learning, Vol. 4 (1989), pp. 227-243.

Morse, D. and W. Shaw, "Investing in Bankrupt Firms," Journal of Finance, Vol. 43 (1988), pp. 1193-1206.

Ohlson, J. A., "Financial Ratios and the Probabilities Prediction of Bankruptcy," Journal of Accounting Research, Vol. 18 (Spring 1980), pp. 109-131.

Quinlan, J. R., "Induction of Decision Trees," Machine Learning, Vol. 1 (1986), pp. 81-106.

Rayburn, J., "The Association of Operating Cash Flow and Accruals with Security Returns," Journal of Accounting Research, Vol. 24 (Supplement 1986), pp. 112-137.

Scott, J., "The Probability of Bankruptcy," Journal of Banking and Finance, Vol. 5 (1981), pp. 317-344.

Shannon, C. E., "A Mathematical Theory of Communication," Bell System Technology Journal, Vol. 27 (July 1948), pp. 379-423, (October 1948), pp. 623-656.

__________, "Prediction and Entropy of Printed English," Bell System Technology Journal, Vol. 30 (January 1951), pp. 50-65.

Shannon, C. E. and W. Weaver, The Mathematical Theory of Communication, The University of Illinois Press, 1963.

Shaw, M. J., J. A. Gentry and S. Piramuthu, "Inductive Learning Methods for Knowledge-Based Decision Support: A Comparative Analysis," Computer Science in Economics and Management, Vol. 3 (1990), pp. 147-165.

Stiglitz, J., "Some Aspects of the Pure Theory of Corporate Finance: Bankruptcies and Takeovers," Bell Journal of Economics and Management Science, (Autumn 1972), pp. .

Tessmer, Antoinette C., "New Dimensions of Inductive Learning for Credit Analysis," Working Paper, College of Commerce and Business Administration, University of Illinois, July 1992, 27 pages.

Warner, J. B., "Bankruptcy, Absolute Priority, and the Pricing of Risky Debt Claims," Journal of Financial Economics, Vol. 4 (1977), pp. 239-276.

White, M. J., "Bankruptcy Costs and the New Bankruptcy Code," Journal of Finance, Vol. 38 (May 1983), pp. 477-488.

Wilson, G. P., "The Incremental Information Content of the Accrual and Funds Components of Earnings After Controlling for Earnings," The Accounting Review, Vol. 62 (April 1987), pp. 293-322.

____________, "The Relative Information Content of Accruals and Cash Flows: Combined Evidence at the Earnings Announcement and Annual Report Release Date," Journal of Accounting Research, Vol. 24 (Supplement 1986), pp. 165-200.

Zmijewski, M. E., "Methodological Issues Related to the Estimation of Financial Distress Prediction Models," Journal of Accounting Research, Vol. 22 (Supplement 1984), pp. 59-82.

EXHIBIT 1

AN EXAMPLE OF CASH FLOW COMPONENTS (CFC)

CASH INFLOWS (+)		CASH OUTFLOWS (-)	
NET OPERATING	\$1220	Δ RECEIVABLES	\$440
Δ OTHER C.A.	40	Δ INVENTORY	360
Δ PAYABLES	200	FIXED COVERAGE EXP.	180
Δ OTHER C.L.	100	NET INVESTMENT	720
Δ NET FINANCIAL	340	DIVIDENDS	300
Δ CASH M.S.	140	Δ NET OTHER A & L	40
TOTAL CASH FLOW (+)	\$2040	TOTAL CASH FLOW (-)	\$2040

AN EXAMPLE OF RELATIVE CASH FLOW COMPONENTS (CFC*)[1]

CASH INFLOWS (+)	% OF TOTAL CASH FLOW (+)	CASH OUTFLOWS (-)	% OF TOTAL CASH FLOW (-)
NET OPERATING*	59.8	Δ RECEIVABLES*	21.6
Δ OTHER C.A.*	2.0	Δ INVENTORY*	17.6
Δ PAYABLES*	9.8	FIXED COVERAGE EXP.*	8.8
Δ OTHER C.L.*	4.9	NET INVESTMENT*	35.3
Δ NET FINANCING*	16.7	DIVIDENDS*	14.7
Δ CASH M.S.*	6.8	Δ NET OTHER A & L*	2.0
	100%		100%

---

1 $\frac{\text{CASH FLOW COMPONENT}}{\text{TOTAL CASH FLOW}}$ = RELATIVE CASH FLOW COMPONENT

*Indicates relative cash flow as opposed to actual cash flow.

EXHIBIT 2

AN EXAMPLE OF THE HIERARCHY OF RELATIVE CASH FLOW COMPONENTS UNDER VARIOUS RISK CONDITIONS

	Company			
	Lowest Credit Risk			Highest Credit Risk
Relative Cash Flow Components (CFC*)	A	B	C	D
Net Operating (NOF*)	92%	70%	57%	15%
ΔAR*	-9	-15	-22	30
ΔINV*	-11	-17	-18	25
ΔOCA*	-1	-3	2	10
ΔAP*	7	15	17	-43
ΔOCL*	1	8	9	-25
Net Investment (NIF*)	-45	-38	-30	-15
Surplus or Deficit after Investment Expenditures	34	20	15	-3
Fixed Coverage Exp. (FCE*)	-2	-6	-9	-16
Surplus or Deficit available for dividends	32	14	6	-19
Dividends (DIV*)	-12	-14	-15	-1
Net Cash Flow Surplus or Deficit (NCF*)	20%	0%	-9%	-20%
ΔNet Financing (ΔNFF*)	-10	7	10	19
ΔNet Other A & L (ΔNOA&L*)	0	0	-6	1
ΔCash & M.S. (ΔCash*)	-10	-7	5	0
CFC* After All Cash Flows	0	0	0	0

EXHIBIT 3

MEANS AND STANDARD DEVIATIONS OF THE RELATIVE CASH FLOW COMPONENTS (CFC*) FOR THE SAMPLE FAILED AND NONFAILED COMPANIES

	Failed Companies		Nonfailed Companies	
CFC* Titles	Mean	S.D.	Mean	S.D.
Operating (NOF*)	.1573	.4686	.4865	.2713
Investment (NIF*)	-.1167	.3220	-.3386	.2484
Dividend (DIV*)	-.0223	.0707	-.0813	.0705
Fixed Coverage (FCE*)	-.1513	.1315	-.1266	.0832
Receivables (ΔARF*)	.0316	.2417	-.0562	.1589
Inventories (ΔINVF*)	-.0025	.2374	-.0587	.1780
Other CA (ΔOCAF*)	.0148	.1358	-.0038	.1012
Payables (ΔAPF*)	.0059	.1920	.0413	.1205
Other CL (ΔOCLF*)	.0206	.1285	.0139	.1188
Other A & L (ΔNOA&LF*)	.0164	.2357	.0067	.1505
Financing (ΔNFF*)	.0645	.3334	.1191	.3138
Change in Cash (ΔCash)	-.0006	.2198	-.0072	.2351
Other Variables				
TCF/TA	.3789	.2783	.2772	.1531
AD/FA[1]	.4122	.1827	.3802	.1608
Sales Trend	.4444	.4969	.2222	.4157
N	99		99	

---

[1]Accumulated Depreciation/Fixed Assets.

FIGURE 1
An Example of Entropy

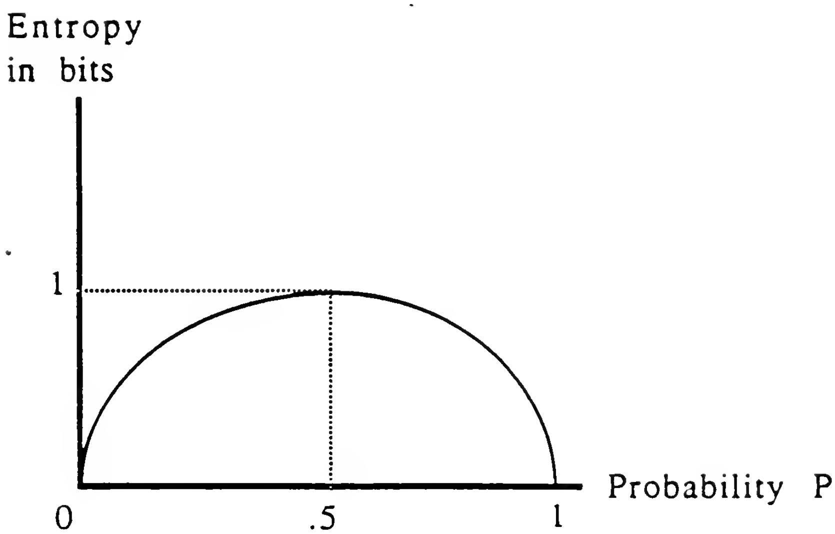
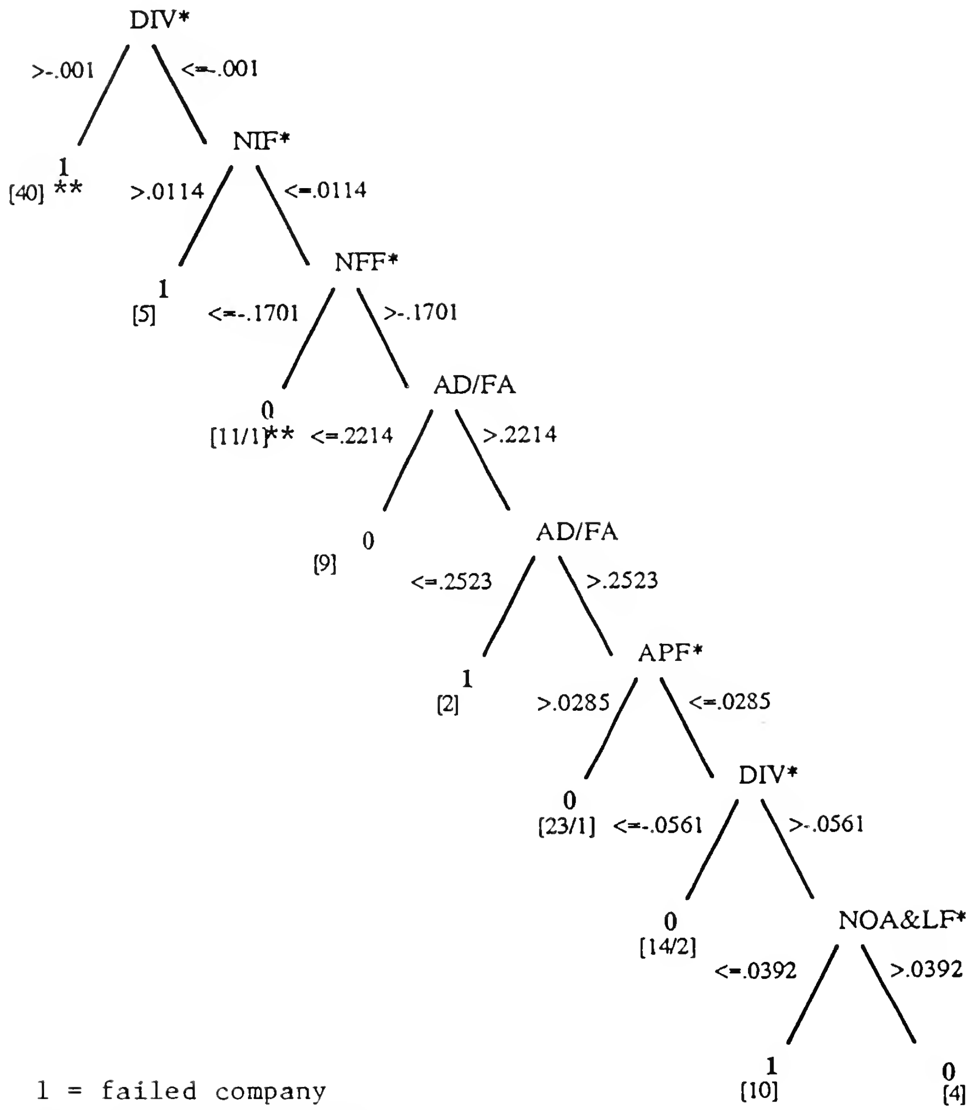


Figure 2

Inductive Learning Tree Based on a Training Sample of 118 Companies

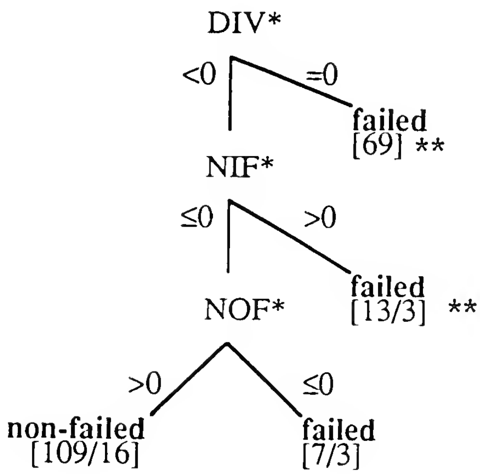


1 = failed company
0 = non-failed company
classification accuracy 96.4%
prediction accuracy 90.05 %

** (n/m) = a total of n companies reach the node,
m of them are misclassified by the node.

Figure 3

A Global Tree of the 198 Companies



** (n/m) = a total of n companies reach the node,
m of them are misclassified by the node.

FIGURE 4

Frequency of DIV*

Bankrupt Firms

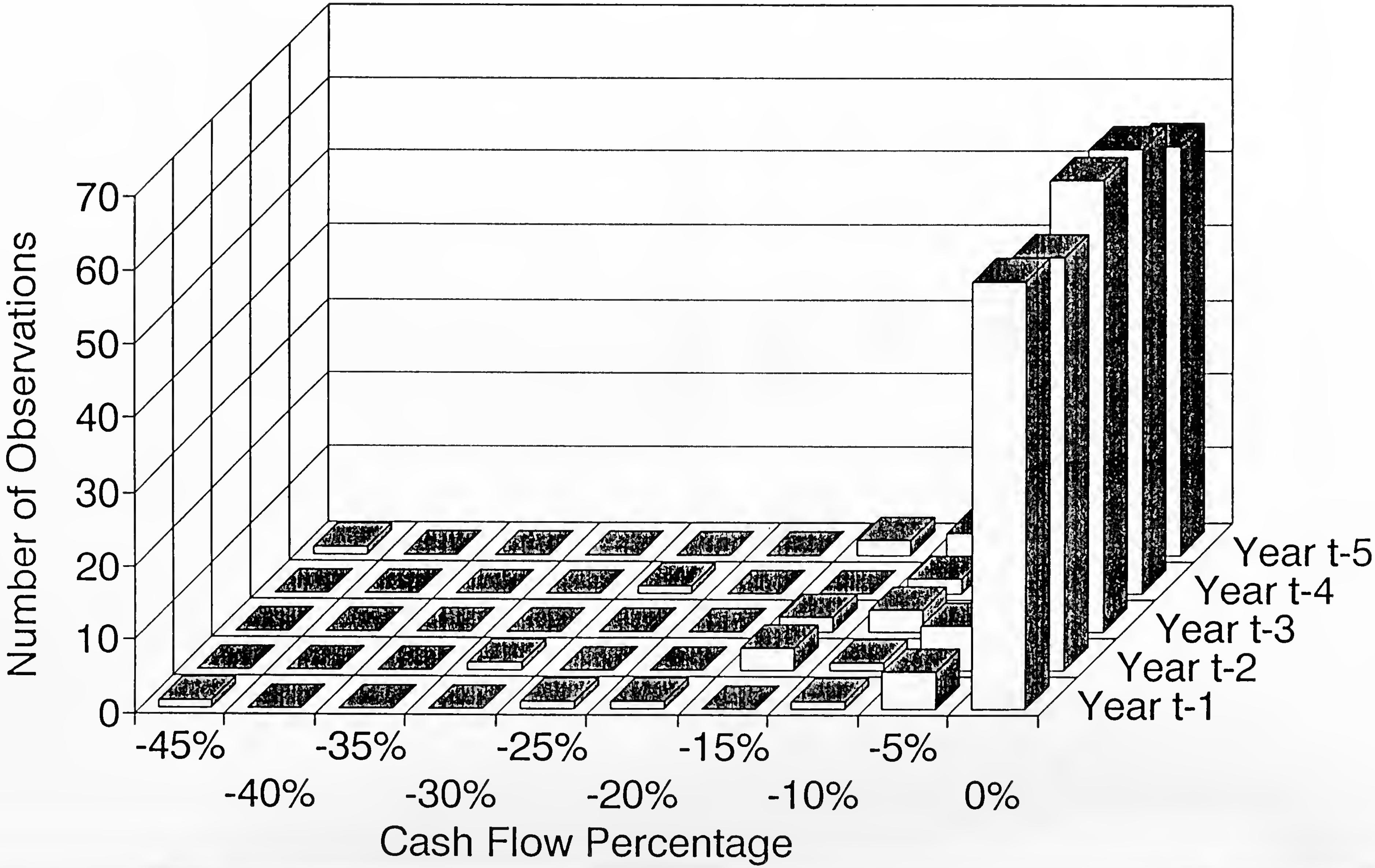
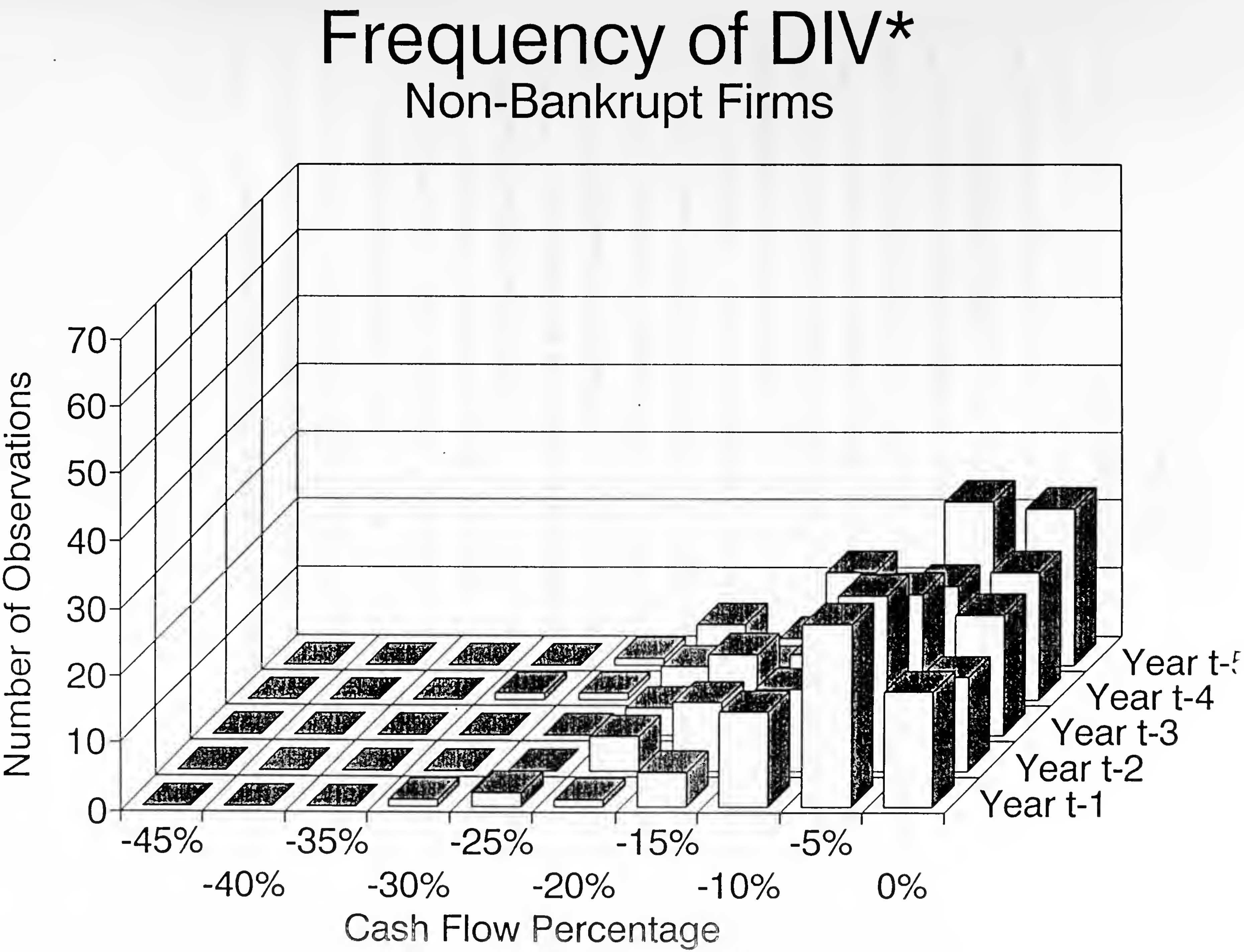


FIGURE 5


Frequency of DIV*
Non-Bankrupt Firms
Number of Observations
70
60
50
40
30
20
10
0
-45%
-40%
-35%
-30%
-25%
-20%
-15%
-10%
-5%
0%
Cash Flow Percentage
Year t-1
Year t-2
Year t-3
Year t-4

APPENDIX A

MEASURING ENTROPY

A simplified training sample of 10 failed and nonfailed companies is presented in Exhibit 4. It is used to illustrate the operation of the ID3 algorithms. Only two classes are used in order to simplify the example. Three attributes are selected among the most important relative cash flow components--net operating (NOF^*), net investment (NIF^*) and dividends (DIV^*). The values for these attributes are found in Exhibit 4. The failed or nonfailed classification may be regarded as Shannon's incoming message to be reproduced as exactly as possible in a decision tree. The classification observed at the final nodes of the tree may be regarded as Shannon's outgoing message, the decision tree being regarded as Shannon's channel for transmitting information.

In this example, there are six nonfailed and four failed companies. The probabilities of failure or nonfailure can be estimated by using the relative frequencies observed in the training sample. If p is the probability of occurrence of nonfailure, then p = 0.6 and the probability of failure is 1 - p = 0.4. The simplest decision tree to reproduce such a message is shown in Figure 6.

The entropy (H) contained in the outgoing message in Figure 6 is the same as the uncertainty contained in the incoming message:

$$H = -0.6 \log_2 0.6 - 0.4 \log_2 0.4 = 0.97. \qquad (1)$$

In other words, the decision tree in Figure 6 does not reduce the uncertainty from incoming to outgoing messages, nor is any information gained.

To improve the decision tree, each attribute (variable) must be evaluated as to its appropriateness to reduce entropy. First, the relative cash outflow going to dividends (DIV^*) is tested, as shown in Figure 7. The data are based on the training sample in Exhibit 4. When DIV^* is low, the amount of entropy contained in the outgoing message of the subtree is

$$-0.6 \log_2 0.6 - 0.4 \log_2 0.4 = 0.97. \qquad (2)$$

When DIV^* is high, the entropy associated with the subtree is also 0.97. Therefore, if the tree is built on DIV^*, the entropy of the outgoing message transmitted by the tree is

$$0.5 * 0.97 + 0.5 * 0.97 = 0.97. \qquad (3)$$

Hence, the amount of information gained by splitting on DIV^*, which is the reduction in entropy by the split, is calculated as the difference between the entropy contained in the simplest tree (H) and the total entropy on DIV^*:

$$0.97 - 0.97 = 0. \qquad (4)$$

In essence, a tree built on DIV^* does not help to gain information.

The second variable to be tested is the relative net operating cash flow (NOF^*), which is shown in Figure 8. When NOF^* is small, the amount of entropy contained in the outgoing message of the subtree is

$$-0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1.0. \tag{5}$$

When NOF^* is medium or large, the entropy associated with both subtrees is zero, which implies that there is no uncertainty. Thus, the expected total entropy after splitting on NOF^* is

$$0.4 * 1.0 + 0.2 * 0 + 0.4 * 0 = 0.4. \tag{6}$$

Therefore, the amount of information gained by using NOF^* as a node is

$$0.97 - 0.40000 = 0.57. \tag{7}$$

The third variable to be tested is relative net investment (NIF^*) which is shown in Figure 9. When NIF^* is low, the entropy contained in the outgoing message transmitted by the subtree is

$$-0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1.0. \tag{8}$$

When NIF^* is high, the entropy associated with the subtree is

$$-0.75 \log_2 0.75 - 0.25 \log_2 0.25 = 0.81. \qquad (9)$$

Thus, the total entropy contained in the outgoing message after splitting on NIF* is

$$0.6 * 1.0 + 0.4 * 0.81 = 0.92. \qquad (10)$$

Hence, the amount of information gained by using NIF* as a node is

$$0.97 - 0.92 = .05. \qquad (11)$$

The largest amount of information gain is obtained by using NOF*. In other words, NOF* provides the largest reduction of uncertainty with respect to analyzing financial failure. Hence, NOF* is chosen as the root node of the tree. If NOF* is used as the root node, there still remains uncertainty (entropy = 0.40) only when NOF* is small. Again, NOF* and DIV* are tested by the same procedure as potential subsequent nodes. Figure 10 shows that when DIV* is low the entropy contained in the outgoing message transmitted by the subtree is

$$-0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1.0. \qquad (12)$$

When DIV* is high, the same entropy is obtained. Thus, the total entropy contained in the outgoing message after splitting on NOF* and DIV* is

$$0.4[0.5 * 1.0 + 0.5 * 1.0] = 0.4. \tag{13}$$

Hence, the amount of information gained by using DIV* as second node is

$$0.4 - 0.4 = 0. \tag{14}$$

which means that DIV* does not help to gain information.

NIF* is then tested as subsequent node which is shown in Figure 11. As all the companies belong to a single class whenever NIF* is low or high, the entropy contained in the outgoing message transmitted by the subtree is

$$-0.\ \log_2 0.\ - 1 \log_2 1 = 0. \tag{15}$$

Thus the total entropy contained in the outgoing message after splitting on NOF* and NIF* is

$$0.4[0.5 * 0. + 0.5 * 0.] = 0. \tag{16}$$

Hence, the amount of information gained by using NIF* as second node is

$$0.4 - 0. = 0.4. \tag{17}$$

Therefore, NIF* is selected as second node and there remains no uncertainty about the outgoing message (entropy = 0.). The inductive process is terminated and Figure 12 shows the final tree.

EXHIBIT 4

FINANCIAL FAILURE TRAINING EXAMPLE

Relative Cash Flow Components

Firm	Investment (NIF*)	Operating (NOF*)	Dividend (DIV*)	Failed or Nonfailed Classifications
A	low	small	low	Failed
B	high	small	low	Nonfailed
C	high	medium	low	Failed
D	low	large	low	Nonfailed
E	high	large	low	Nonfailed
F	low	small	high	Failed
G	high	large	high	Nonfailed
H	high	small	high	Nonfailed
I	low	medium	high	Failed
J	low	large	high	Nonfailed

FIGURE 6

Initial Decision Tree

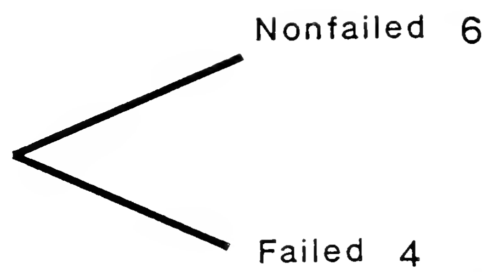


FIGURE 7

DIV* Decision Tree

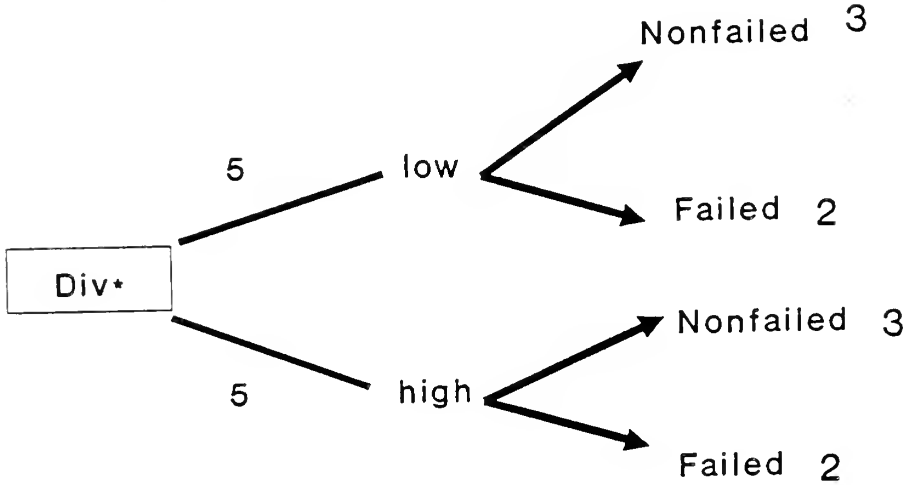


FIGURE 8
NOF* Decision Tree

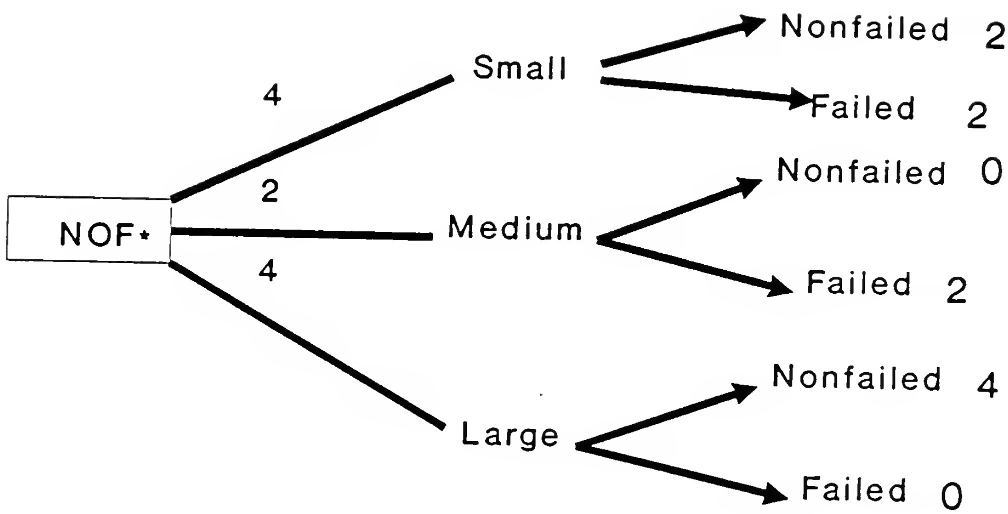


FIGURE 9
NIF* Decision Tree

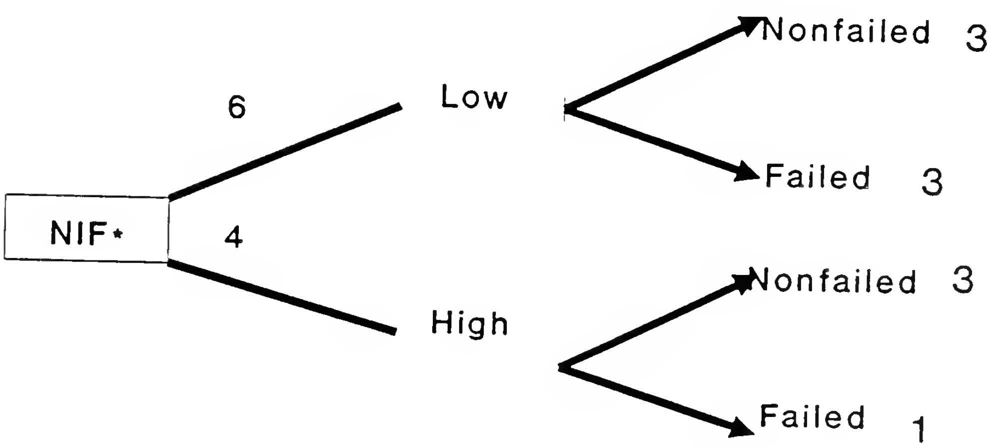


FIGURE 10
NOF* and DIV* Decision Tree

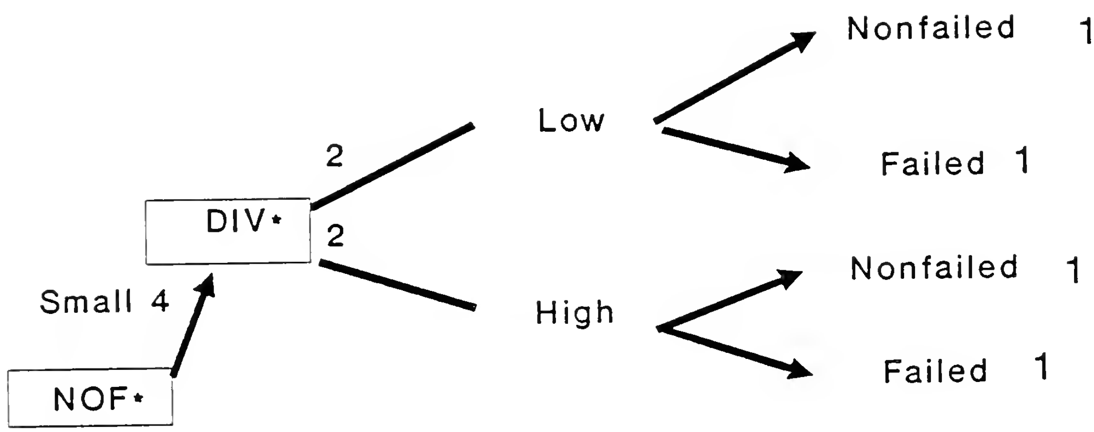


FIGURE 11
NOF* and NIF* Decision Tree

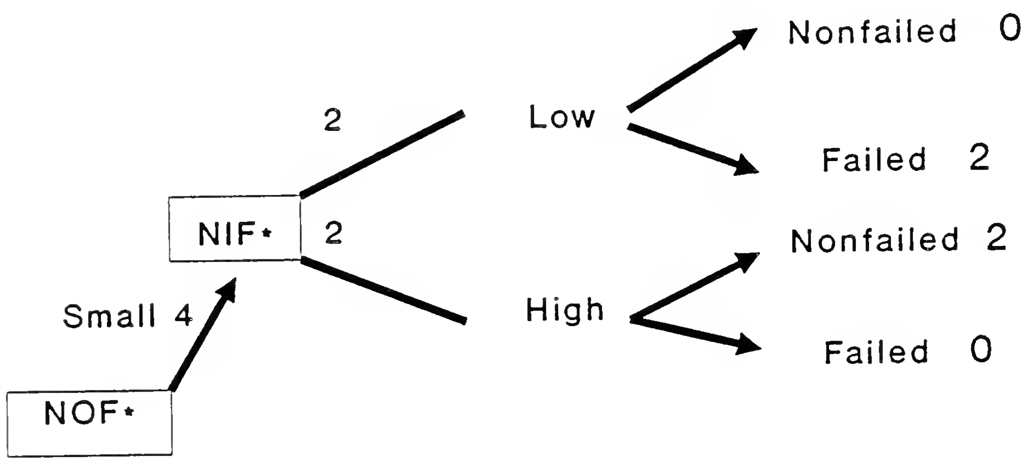


FIGURE 12

Final Decision Tree